

Amendments to the Claims

Claim 1 (currently amended). A system for disambiguating speech input using multimodal interaction with an application comprising:

a speech recognition component that receives recorded audio or speech input and generates:

one or more tokens corresponding to the speech input; and

for each of the one or more tokens, a confidence value indicative of the likelihood that the a given token correctly represents the speech input;

a selection component that identifies, according to a selection algorithm, which two or more tokens are to be presented to a user as alternatives, wherein said alternatives are words or tokens;

one or more disambiguation components that perform an said multimodal interaction with the user to present the alternatives to the user and to receive a selection of alternatives from the user, the interaction taking place in at least a visual mode wherein the multimodal interaction allows input and output in voice and visual modes; and

an output interface that presents the selected alternative to an application as input.

Claim 2 (original). The system of claim 1, wherein the disambiguation components and the application reside on a single computing device.

Claim 3 (original). The system of claim 1, wherein the disambiguation components and the application reside on separate computing devices.

Claim 4 (original). The system of claim 1, wherein the one or more disambiguation components perform said interaction by presenting the user with alternatives in a visual mode, and by receiving the user's selection in a visual mode.

Claim 5 (original). The system of claim 4, wherein the disambiguation components present the alternatives to the user in a visual form and allow the user to select from among the alternatives using a voice input.

Claim 6 (original). The system of claim 1, wherein the one or more disambiguation components perform said interaction by presenting the user with alternatives in a visual mode, and by receiving the user's selection either in a visual mode, a voice mode, or a combination of visual mode and voice mode.

Claim 7 (original). The system of claim 1, wherein the selection component filters the one or more tokens according to a set of parameters.

Claim 8 (original). The system of claim 7, wherein the set of parameters is user specified.

Claim 9 (original). The system of claim 1, wherein the one or more disambiguation components disambiguates the alternatives in plural iterative stages, whereby the first stage narrows the alternatives to a number of alternatives that is smaller than that initially generated by the selection component, but greater than one, and whereby the one or more disambiguation components operative iteratively to narrow the alternatives in subsequent iterative stages.

Claim 10 (original). The system of claim 9, whereby the number of iterative stages is limited to a specified number.

Claim 11 (currently amended). A method of processing speech input using multimodal interaction with an application comprising:

receiving a speech input from a user;

determining whether the speech input is ambiguous;

if the speech input is not ambiguous, then communicating a token representative of the speech input to an application as input to the application; and

if the speech input is ambiguous;

performing an a multimodal interaction with the user whereby the user is presented with plural alternatives and selects an alternative from among the plural alternatives, wherein said alternatives are words or tokens, the interaction being performed in at least a visual mode wherein said multimodal interaction allows input and output in voice and visual modes;

communicating the selected alternative to the application as input to the application.

Claim 12 (original). The method of claim 11, where the interaction comprises the concurrent use of said visual mode and said voice mode.

Claim 13 (original). The method of claim 12, wherein the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input.

Claim 14 (original). The method of claim 11, wherein the interaction comprises the user selecting from among the plural alternatives using visual input.